DOCUMENT RESUME

ED 296 813 RC 016 589

AUTHOR Brescia, William; Fortune, Jim C.
TITLE Standardized Testing of American Indian Students. ERIC Digest.
INSTITUTION ERIC Clearinghouse on Rural Education and Small Schools, Las Cruces, N. Mex.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO EDO-RC-88-15
PUB DATE Mar 88
CONTRACT 400-86-0024
NOTE 13p.; Because this Digest was published in colored ink, on colored paper, and in reduced type size (for compression), a full-size typescript version has also been included here.
PUB TYPE Guides - Non-Classroom Use (055) -- Information Analyses - ERIC Information Analysis Products (071)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Achievement Tests; *American Indian Education; American Indians; Aptitude Tests; Cultural Differences; *Culture Fair Tests; Elementary Secondary Education; *Standardized Tests; Student Characteristics; Test Bias; Testing; *Testing Problems; Test Interpretation; *Test Validity
IDENTIFIERS ERIC Digests

ABSTRACT

The testing of many American Indian children using standardized exams (achievement, aptitude, ability, and intelligence tests) developed for the majority American society represents a case of cross-cultural testing which is likely to produce invalid results in the form of underestimation of student performance. Requirements for obtaining accurate and appropriate scores include matched level of test to the students, clarity and comprehensiveness in the administration of instructions, and logistics. Requirements associated with test interpretation include norms, reliability, and validity. Underestimation may occur in the standardized testing of American Indian students by students not exhibiting behaviors required in successful test-taking, students not reading the questions accurately, students not having the assumed experience or cognitive structure to respond to certain items, and students lacking the opportunity to practice key behaviors required by the test. Many American Indian students fail to exhibit successful test-taking behaviors due to a multiplicity of underlying causes: cultural beliefs, language, and the setting in which many children are reared. Test users can contribute to better test-taker performance by assuring that students have had exposure to experiences assumed in the test design, opportunities to develop requisite skills, circumstances necessary to value successful test performance, and an understanding of instructions and time requirements. (NEC)

ED296813

ERIC DIGEST

AMERICAN INDIAN EDUCATION

EDO-RC-88-15

Standardized Testing of American Indian Students

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Introduction

Testing students from backgrounds different from the culture in which the test was developed magnifies the probability of invalid results. In addition to the limits of test theory and the constraints associated with a given test, the test administrator is faced with several potential sources of error arising from the differences in the two cultures. These include lack of compatibility of the languages, differences in the experiential backgrounds of the students being tested and those for whom the test was developed, and differences in affective dispositions toward handling testing environments between the two groups of students.

The testing of many American Indian children using exams developed for the majority American society represents a case of cross-cultural testing which is likely to produce invalid results in the form of underestimation of student performance. In this digest limitations of the use of standardized tests with many American Indian students are discussed as a problem that must be addressed by educators who use these tests for placement and grading of students. Sources and consequences of invalid test results ensuing from the administration of standardized tests to unacculturated American Indian students are delineated.

Overview of Standardized Testing

What is a standardized test?

Technically, a standardized test is a systematic process for assigning numerical values to samples of behavior, such that the same or equivalent items are administered to all test-takers using uniform directions and scoring methods (Brown, 1976). Standardization can be extended to include uniform methods of test interpretation derived through the development of norms for the group for which the test was designed, information on the consistency of test performance (reliability), and evidences that the test measures what it was designed to measure (validity) (Aiken, 1976).

Tests addressed in this digest will be limited to four types which are commercially available. These include achievement, aptitude, ability, and intelligence tests. These tests are those that are most frequently used in educational placement and grading.

RC 016 589

Achievement Tests. An achievement test is one that has been designed to measure the knowledge and skills accrued in a specific content area. The content areas are usually school subjects taught at a given grade level. Commercially available achievement tests are most frequently multiple choice, group-administered tests of content selected to apply to a wide range of school programs.

Aptitude tests. An aptitude test is a test that has been designed to measure the capability of an individual to profit from instruction in a specific content area. An aptitude test is designed to measure skills, traits, and talents predictive of future performance in the area.

Ability tests. An ability test is one designed to measure the capability of one to perform in a content area. The difference between an ability test and an aptitude test is one of status; present capability versus potential (Brown, 1976).

Intelligence tests. An intelligence test is one designed to measure an individual's ability to reason and to perform verbally. The IQ test is a generalized form of aptitude test for scholastic work (Aiken, 1976).

Performance Criteria for Standardized Tests

Criteria for the performance of a standardized test can be classified into two categories: requirements for obtaining accurate and appropriate scores; and requirements for making accurate and useful interpretations of the scores. Requirements for obtaining accurate and appropriate scores include matched level of test to the students, clarity and comprehensiveness in the administration of instructions, and logistics. Requirements associated with test interpretation include norms, reliability, and validity.

Level of Test. Tests are designed to measure specific knowledge, skills, and abilities of students sharing a particular array of characteristics. The test will likely not perform for students who do not have the characteristics. Hence, tests should be used with students for whom test norms exist.

Clarity and Comprehensiveness of Instructions. Noise during testing, distractions, incomplete instructions, failure to allow the time allotted to the test (mistiming), and inadequate coverage of practice examples can reduce

student performances. Instructions should be brief and clear. They should also show what the test-taker should do to respond, and tell that individual how much time is available for each part of the test.

Logistics. Tests should be administered under favorable conditions and should be monitored. Rushed, disorderly arrangements raise the potential of test anxiety. Cheating either from copying or from obtaining prior copies of the test can invalidate any test. Harshness or condescending remarks can lower the performance of a student on a test.

Norms. Norms permit an interpretation of a test taker's score to be made relative to the scores made by a large number of similar individuals. These interpretations do not address how much content was mastered but how well the test-taker did in comparison to his/her peers. In the case of achievement testing, test administration should be in the same time of the year as the test was normed. Most achievement tests have fall and spring norms. When tests are administered to members of a minority group, it is very important to have norms on the minority group or to assure that the norm group includes adequate representation of the minority group.

Reliability. Reliability addresses the reproducibility of test performances across time, as well as forms of the test and items in it. Low reliability is due to unsystematic errors which affect test scores in unpredictable ways. These unsystematic errors may occur due to ambiguous items, guessing, or poor content coverage. Poor content coverage may involve uneven, sparse, or spotty coverage of content. Standardized tests used for educational decisions concerning individual students should have reliabilities greater than 90 percent.

Validity. Uses of a test are subject to validity concerns. One test may be valid as a measure of the achievement of a mathematics unit within the context it was taught and yet not be a valid measure for the generic achievement of the mathematics included in the unit. If addition of several numbers is taught vertically and all of the test items are arranged vertically then the test may be a valid measure of the unit within the instructional context. Yet, the test is not a valid measure of the addition of several numbers, since the numbers can also be arranged horizontally. Validity refers to measuring what the use is referencing.

In tests designed to measure acquisition of a body of content, such as achievement tests, content validity is a required criterion. It is of vital importance that the designated content is covered adequately and in the appropriate proportions. Administrative manuals of tests usually describe the table of content specification as evidence of content validity. In achievement testing, instructional validity or evidence that the content was taught is also important.

In tests designed to measure potential, predictive validity is required. Predictive validity is evidence that shows that performance on the test is correlated with future performance in the content area. A second validity concern with tests that measure potential is that of test bias. Test bias occurs when individuals with a given characteristic not logically related to skill consistently score differently than the norm group on the test.

Application to American Indian Education

Use of Standardized Tests

Standardized test scores are used to make decisions about programs and to make decisions about individuals. Although all four types of tests can be used indirectly to make decisions about programs, the primary uses of tests to make program decisions involve the use of ability tests in program design and the use of achievement tests in program evaluation.

All four types of tests are used to make decisions about individuals. Aptitude, ability, and intelligence tests are used primarily in the rendering of decisions involving selection and placement or to provide feedback to the test-taker concerning his or her capability. Achievement tests may be used in four ways to make decisions about individuals as a survey of attainment in a content area, as a diagnostic instrument to identify the strengths and weaknesses of an individual in a content area, as a readiness indicator to determine if an individual has attained enough prerequisite material to continue study in a given content area, and as a performance test to estimate the degree of learning of a body of content defined by instruction.

Limitations to Use

Generally, when standardized tests are used with American Indian students (on the reservation or in settings with low levels of acculturation) and produce invalid results, the tests usually produce lower or less desirable scores for the Indian test-taker. These score variations are not really explained by program related factors nor correlates of test performance which are frequently found in other situations.

In program-related decisions the underestimation of Indian performance on ability tests may result in the development of an inefficient program design. Underestimation on achievement tests may result in the demise or modification of what in reality is an effective program.

In test-based decisions concerning the American Indian student underestimation can do grave harm to the individual. In both selection and placement the Indian student can be denied opportunity, can be relegated to low-paying work or unemployment or can be placed in a program that is too easy or boring. As feedback, the test results can do harm to the self esteem and confidence of the Indian student, sometimes resulting in the student giving up or dropping out. With regard to the interpretation of achievement test results, false conclusions concerning the Indian student may result leading to teacher allegations of laziness, disinterest, or stupidity. Underestimation may result in the student not being promoted to advanced levels of instruction, being placed in low-achieving groups, or having to do remedial work.

Sources of Test Underestimation

It should be recognized that American Indian tribes

embrace a wide range of differences with regard to culture. To treat American Indians as a collective group, regardless of tribe, is to make the same error of consideration as is made in testing Indian students with standardized tests that have less than three percent Indian students in the norming sample. Uniform research results across tribes is simple nonexistent. Nevertheless, bias found in the study of one tribe likely exists for several other tribes. A source of underestimation documented for one tribe should in fact be considered as a potential source of underestimation for other tribes until research indicates the contrary for a given tribe.

Underestimation may occur in the standardized testing of American Indian students in several different ways. These include students not exhibiting behaviors required in successful test-taking; students not reading the questions accurately; students not having the assumed experience or cognitive structure to respond to certain items; and students lacking the opportunity to practice key behaviors required by the test. Each of these differences in the behavior of Indian students in the testing situation reflects cultural differences.

The factors that influence Indian test scores, usually considered as forms of bias, are well-documented in the literature (Neely and Shaughnessy, 1984). If only one of the unsuccessful test-taking behaviors of the students is present in certain situations or if these behaviors could be tracked systematically, then a methodology could be developed to correct the problem of underestimation. However, these behaviors are confounded in that they sometimes occur jointly and at different times in the test-taking process. Additional confounding takes place because many Indian students possess the other individual characteristics which normally present testing problems: poverty, low parental education, broken homes, and nonstandard English backgrounds. McDiarmid (1972) discusses the role that poverty, health and nutrition, social conflict, language, and test motivation play in the interpretation of test data on Indian children. The major factors were found to be language and test motivation. Some suggestions to facilitate test fairness have been reported in aptitude and ability assessment, such as in the General Aptitude Test Battery (Hunter, 1983). Measurement professionals have addressed the problem of cultural influence on test performance, but to this date an operationally functional treatment of the problem still does not exist.

Many American Indian students fail to exhibit successful test-taking behaviors due to a multiplicity of underlying causes. Cultural beliefs in some tribes may bar competitive behaviors in an academic setting. The student may underestimate the seriousness of the test or fail to adopt a successful response strategy which may involve selective scanning for known items, techniques of using partial information to guess correct answers, or efficient time use. Students possess a dichotomy with regard to their perceptions of the purposes and significance of tests (Deyhle, 1986). Acculturation has been found as an influence on both achievement and ability tests (Hoffman, et al, 1985; Guilmet, 1983). Both authors suggest that acculturation and test motivation are associated.

The second most influential factor reported in the literature is language; that is, inability of many Indian students to read the questions accurately or to be able to give appropriate verbal responses. Tests which avoid language use are not subject to underestimation as much as those that depend on verbal instructions and reading. To illustrate this the Hiskey-Nebraska test of Learning Aptitude, a non-verbal test designed for use with deaf children produced estimates of higher potential for Indian children than the Wechsler Test (Shutt, 1962). The influence of English being learned as a second language is further reinforced by the fact that many Indian first languages are unwritten.

The deficiency of the student to have the assumed experience or cognitive structure to respond to certain items is due both to the culture and to the setting in which many children are reared. The isolated, rural environment of many reservation settings, the restricted poverty of many families, and the cultural ties that promote continued identification with the tribe deny students important knowledge of the outside world. Fortune (1985) found that a majority of the Indian students in an economically deprived reservation setting lacked experience to understand the examples that teachers use in instruction and, consequently, the experience to perform on achievement tests. Studies of intelligence and aptitude test results of one tribe suggest nonstandard Indian performances (Mishra, 1981). These results are further substantiated by unique patterns of measured Indian aptitude. On the WISC-R test Indian children show a pattern of greater strength than the norm population in relational, holistic, and right hemisphere information processing (Browne, 1984). The Indian children in two reservation settings demonstrated a performance pattern on the Wechsler different from normal and learning disabled Anglo children. Spatial abilities were more well-developed than sequencing skills (McShane & Plas, 1982; Diessner & Walker, 1986).

Although Boloz and Varrati (1983) found that test scores for Indian students were higher for those who had the best attendance records and stayed in the same district, many Indian students live in discouraging situations where there is little congruence between their life experiences and the skills needed for testing. These students often do not speak English outside of school. In addition, there are few books available for them to read. Personal and community poverty, aggravated by lack of industrial development and employed role models, does little to stimulate student awareness of mathematics and its applications.

Considerations for Test Use

Applications of the principles of test theory to the sources of underestimation can lead to several useful maxims in using standardized tests with many American Indian students. Prior to the administration of standardized tests to American Indian students or to the interpretation of American Indian test data the test user can do several things that may contribute to better test-taker performance. For ability, aptitude, and intelli-

gence testing one should assure that the students have had exposure to the experiences assumed in the design of the test, the opportunity to develop the requisite skills, and the circumstances necessary to value a successful test performance. For achievement tests, one should make certain that the students have been instructionally exposed to the content of the test and have had opportunity to apply this content; the students have had experience in the taking of the test, are test wise and able to understand test instructions and time requirements; and the test is to be or was administered at a time similar to when it was normed and the test has Indian norms. A few of the national testing corporations such as the developers of the California Achievement Test are developing Indian norms for their tests.

Several papers offer additional reading and help in the area of testing Indian students. They include guidelines for testing bicultural children (Bernardoni, 1967) and for second language testing (Upshur and Fata, 1968), as well as annotations of tests found appropriate for use with American Indians (Educational Testing Service, 1977 and 1982).

Recommendations for future research appear fraught with problems. The natural recommendation for most tribes to become involved in the process of developing their own tests has to be considered in the light of the high costs and resource requirements needed to develop a quality test. Tribal differences and dispersion of many Indian students would suggest that tribally developed tests may lack enough general applications across tribes for merit. Further research may be better invested in documenting the similarity and differences of test reactions and in the development of intervention programs to teach test administrators to use tests in an appropriate manner with American Indian students.

REFERENCES

Aiken, Lewis R. PSYCHOLOGICAL TESTING AND ASSESSMENT. Boston: Allyn and Bacon, Inc., 1976.

Bernardoni, Louis C. "The Testing of Bicultural Children", SHARING IDEAS 4, 2 (1967): 1-5. ED 077 977.

Boloz, Sigmund A., & Varrati, Richard. "Apologize or Analyze: Measuring Academic Achievement in the Reservation School," JOURNAL OF AMERICAN INDIAN EDUCATION, 23, 1 (1983): 23-8.

Brown, Frederick G. PRINCIPLES OF EDUCATIONAL AND PSYCHOLOGICAL TESTING. New York: Holt, Rinehart and Winston, 1976.

Browne, Dauna Bell. "WISC-R Scoring Patterns among Native Americans of the Northern Plains," WHITE CLOUD JOURNAL 3, 2 (1984): 3-16.

Deyhle, Donna. "Success and Failure: A Micro-ethnographic Comparison of Navajo and Anglo Students' Perceptions of Testing," CURRICULUM INQUIRY, 16, 4 (1986): 365-89.

Diessner, Rhett, & Walker, Jacqueline L. "A Cognitive Pattern of the Yakima Indian Students," JOURNAL OF AMERICAN INDIAN EDUCATION 25, 2 (1986) 39-43.

Educational Testing Service (ETS). TESTS FOR AMERICAN INDIANS. Princeton: ETS, 1977. ED 213 546.

Educational Testing Service (ETS). TESTS FOR AMERICAN INDIANS. Princeton: ETS, 1982. ED 227 995.

Fortune, Jim C. CHOCTAW COMPREHENSIVE SCHOOL STUDY. Philadelphia, MS: Choctaw Heritage Press, 1985.

Guilmet, George M. THE INAPPROPRIATENESS OF STANDARDIZED TESTING IN A CULTURALLY HETEROGENEOUS MILIEU: A NAVAJO EXAMPLE. Los Angeles: University of California, 1983. ED 261 830.

Hoffman, Tom, et al. "Measured Acculturation and MMPI-168 Performance of Adults," JOURNAL OF CROSS-CULTURAL PSYCHOLOGY 16, 2 (1985): 243-56.

Hunter, John E. FAIRNESS OF THE GENERAL APTITUDE TEST BATTERY: ABILITY DIFFERENCES AND THEIR IMPACT ON MINORITY HIRING RATES. Uses Test Research Report No. 46. Sacramento: California State Department of Employment Development, 1983. ED 237 534.

McDiarmid, G. L. THE HAZARDS OF TESTING INDIAN CHILDREN. Las Cruces: ERIC/CRESS, 1972. ED 055 692.

McShane, Damian Anthony, & Plas, Jeanne M. "Wechsler Scale Performance Patterns of American Indian Children." PSYCHOLOGY IN THE SCHOOLS 19, 1 (1982): 8-17.

Mishra, Shitala P. "Relationship of WISC-R Factor Scores to Academic Achievement and Classroom Behaviors of Native American Navajos." MEASUREMENT AND EVALUATION IN GUIDANCE 14, 1 (1981): 26-30.

Neely, Renee, & Shaughnessy, Michael F. ASSESSMENTS AND THE NATIVE AMERICAN. Portales: Eastern New Mexico University Psychology Department, 1984. ED 273 889.

Shutt, Darold L. FAMILY PARTICIPATION IN THE PSYCHOLOGICAL EVALUATION OF MINORITY CHILDREN. Paper presented to Southwestern Orthopsychological Association Meeting, Galveston, TX, November, 1972. ED 071 830.

Upshur, John A., & Fata, Julia, (Eds). PROBLEMS IN FOREIGN LANGUAGE TESTING: PROCEEDINGS OF A CONFERENCE HELD AT THE UNIVERSITY OF MICHIGAN, SEPTEMBER, 1967. Ann Arbor, MI: University of Michigan Research Club in Language Learning, 1968. ED 022 162.

Prepared by:

William Brescia
Mississippi Band of Choctaw Indians
Ph.D. Candidate
University of Indiana
Bloomington, IN

Jim C. Fortune
Virginia Polytechnic Institute
Blacksburg, VA

March 1988

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI) U.S. Department of Education, under contract number 400-86-0024. Its contents do not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. government.

For further information contact: ERIC/CRESS, Dept. 3AP, Box 30001, New Mexico State University, Las Cruces, N.M. 88003-0001 (505) 646-2623

ERIC DIGEST

AMERICAN INDIAN EDUCATION

EDO-RC-88-15

Standardized Testing of American Indian Students

By

William Brescia

and

Jim C. Fortune

March 1988

ERIC® CLEARINGHOUSE ON RURAL EDUCATION and SMALL SCHOOLS

Standardized Testing of American Indian Students

Introduction

Testing students from backgrounds different from the culture in which the test was developed magnifies the probability of invalid results. In addition to the limits of test theory and the constraints associated with a given test, the test administrator is faced with several potential sources of error arising from the differences in the two cultures. These include lack of compatibility of the languages, differences in the experiential backgrounds of the students being tested and those for whom the test was developed, and differences in affective dispositions toward handling testing environments between the two groups of students.

The testing of many American Indian children using exams developed for the majority American society represents a case of cross-cultural testing which is likely to produce invalid results in the form of underestimation of student performance. In this digest limitations of the use of standardized tests with many American Indian students are discussed as a problem that must be addressed by educators who use these tests for placement and grading of students. Sources and consequences of invalid test results ensuing from the administration of standardized tests to unacculturated American Indian students are delineated.

Overview of Standardized Testing

What is a standardized test?

Technically, a standardized test is a systematic process for assigning numerical values to samples of behavior, such that the same or equivalent items are administered to all test-takers using uniform directions and scoring methods (Brown, 1976). Standardization can be extended to include uniform methods of test interpretation derived through the development of norms for the group for which the test was designed, information on the consistency of test performance (reliability), and evidences that the test measures what it was designed to measure (validity) (Aiken, 1976).

Tests addressed in this digest will be limited to four types which are commercially available. These include achievement, aptitude, ability, and intelligence tests. These tests are those that are most frequently used in educational placement and grading.

Achievement Tests. An achievement test is one that has been designed to measure the knowledge and skills accrued in a specific content area. The content areas are usually school subjects taught at a given grade level. Commercially available achievement tests are most frequently multiple choice, group-administered tests of content selected to apply to a wide range of school programs.

Aptitude tests. An aptitude test is a test that has been designed to measure the capability of an individual to profit from instruction in a specific content area. An aptitude test is designed to measure skills, traits, and talents predictive of future performance in the area.

Ability tests. An ability test is one designed to measure the capability of one to perform in a content area. The difference between an ability test and an aptitude test is one of status; present capability versus potential (Brown, 1976).

Intelligence tests. An intelligence test is one designed to measure an individual's ability to reason and to perform verbally. The IQ test is a generalized form of aptitude test for scholastic work (Aiken, 1976).

Performance Criteria for Standardized Tests

Criteria for the performance of a standardized test can be classified into two categories: requirements for obtaining accurate and appropriate scores; and requirements for making accurate and useful interpretations of the scores. Requirements for obtaining accurate and appropriate scores include matched level of test to the students, clarity and comprehensiveness in the administration of instructions, and logistics. Requirements associated with test interpretation include norms, reliability, and validity.

Level of Test. Tests are designed to measure specific knowledge, skills, and abilities of students sharing a particular array of characteristics. The test will likely not perform for students who do not have the characteristics. Hence, tests should be used with students for whom test norms exist.

Clarity and Comprehensiveness of Instructions. Noise during testing, distractions, incomplete instructions, failure to allow the time allotted to the test (mistiming), and inadequate coverage of practice examples can reduce student performances. Instructions should be brief and clear. They should also show what the test-taker should do to respond, and tell that individual how much time is available for each part of the test.

Logistics. Tests should be administered under favorable conditions and should be monitored. Rushed, disorderly arrangements raise the potential of test anxiety. Cheating either from copying or from obtaining prior copies of the test can invalidate any test. Harshness or condescending remarks can lower the performance of a student on a test.

Norms. Norms permit an interpretation of a test taker's score to be made relative to the scores made by a large number of similar individuals. These interpretations do not address how much content was mastered but how well the test-taker did in comparison to his/her peers. In the case of achievement testing, test administration should be in the same time of the year as the test was normed. Most achievement tests have fall and spring norms. When tests are administered to members of a minority group, it is very important to have norms on the minority group or to assure that the norm group includes adequate representation of the minority group.

Reliability. Reliability addresses the reproducibility of test performances across time, as well as forms of the test and items in it. Low reliability is due to unsystematic errors which affect test scores in unpre- dictable ways. These unsystematic errors may occur due to ambiguous items, guessing, or poor content coverage. Poor content coverage may involve uneven, sparse, or spotty coverage of content. Standardized tests used for educational decisions concerning individual students should have reliabilities greater than 90 percent.

Validity. Uses of a test are subject to validity concerns. One test may be valid as a measure of the achievement of a mathematics unit within the context it was taught and yet not be a valid measure for the generic

achievement of the mathematics included in the unit. If addition of several numbers is taught vertically and all of the test items are arranged vertically then the test may be a valid measure of the unit within the instructional context. Yet, the test is not a valid measure of the addition of several numbers, since the numbers can also be arranged horizontally. Validity refers to measuring what the use is referencing.

In tests designed to measure acquisition of a body of content, such as achievement tests, content validity is a required criterion. It is of vital importance that the designated content is covered adequately and in the appropriate proportions. Administrative manuals of tests usually describe the table of content specification as evidence of content validity. In achievement testing, instructional validity or evidence that the content was taught is also important.

In tests designed to measure potential, predictive validity is required. Predictive validity is evidence that shows that performance on the test is correlated with future performance in the content area. A second validity concern with tests that measure potential is that of test bias. Test bias occurs when individuals with a given characteristic not logically related to skill consistently score differently than the norm group on the test.

Application to American Indian Education

Use of Standardized Tests

Standardized test scores are used to make decisions about programs and to make decisions about individuals. Although all four types of tests can be used indirectly to make decisions about programs, the primary uses of tests to make program decisions involve the use of ability tests in program design and the use of achievement tests in program evaluation.

All four types of tests are used to make decisions about individuals. Aptitude, ability, and intelligence tests are used primarily in the rendering of decisions involving selection and placement or to provide feedback to the test-taker concerning his or her capability. Achievement tests may be used in four ways to make decisions about individuals as a survey of attainment in a content area, as a diagnostic instrument to identify the strengths and weaknesses of an individual in a content area, as a readiness indicator to determine if an individual has attained enough prerequisite material to continue study in a given content area, and as a performance test to estimate the degree of learning of a body of content defined by instruction.

Limitations to Use

Generally, when standardized tests are used with American Indian students (on the reservation or in settings with low levels of acculturation) and produce invalid results, the tests usually produce lower or less desirable scores for the Indian test-taker. These score variations are not really explained by program related factors nor correlates of test performance which are frequently found in other situations.

In program-related decisions the underestimation of Indian performance on ability tests may result in the development of an inefficient program design. Underestimation on achievement tests may result in the demise or modification of what in reality is an effective program.

In test-based decisions concerning the American Indian student underestimation can do grave harm to the individual. In both selection and placement the Indian student can be denied opportunity, can be relegated to low-paying work or unemployment or can be placed in a program that is too easy or boring. As feedback, the test results can do harm to the self esteem and confidence of the Indian student, sometimes resulting in the student giving up or dropping out. With regard to the interpretation of achievement test results, false conclusions concerning the Indian student may result leading to teacher allegations of laziness, disinterest, or stupidity. Underestimation may result in the student not being promoted to advanced levels of instruction, being placed in low-achieving groups, or having to do remedial work.

Sources of Test Underestimation

It should be recognized that American Indian tribes embrace a wide range of differences with regard to culture. To treat American Indians as a collective group, regardless of tribe, is to make the same error of consideration as is made in testing Indian students with standardized tests that have less than three percent Indian students in the norming sample. Uniform research results across tribes is simple nonexistent. Nevertheless, bias found in the study of one tribe likely exists for several other tribes. A source of underestimation documented for one tribe should in fact be considered as a potential source of underestimation for other tribes until research indicates the contrary for a given tribe.

Underestimation may occur in the standardized testing of American Indian students in several different ways. These include students not exhibiting behaviors required in successful test-taking; students not reading the questions accurately; students not having the assumed experience or cognitive structure to respond to certain items; and students lacking the opportunity to practice key behaviors required by the test. Each of these differences in the behavior of Indian students in the testing situation reflects cultural differences.

The factors that influence Indian test scores, usually considered as forms of bias, are well-documented in the literature (Neely and Shaughnessy, 1984). If only one of the unsuccessful test-taking behaviors of the students is present in certain situations or if these behaviors could be tracked systematically, then a methodology could be developed to correct the problem of underestimation. However, these behaviors are confounded in that they sometimes occur jointly and at different times in the test-taking process. Additional confounding takes place because many Indian students possess the other individual characteristics which normally present testing problems: poverty, low parental education, broken homes, and nonstandard English backgrounds. McDiarmid (1972) discusses the role that poverty, health and nutrition, social conflict, language, and test motivation play in the interpretation of test data on Indian children. The major factors were found to be language and test motivation. Some suggestions to facilitate test fairness have been reported in aptitude and ability assessment, such as in the General Aptitude Test Battery (Hunter, 1983). Measurement professionals have addressed the problem of cultural influence on test performance, but to this date an operationally functional treatment of the problem still does not exist.

Many American Indian students fail to exhibit successful test-taking behaviors due to a multiplicity of underlying causes. Cultural beliefs in some tribes may bar competitive behaviors in an academic setting. The student may underestimate the seriousness of the test or fail to adopt a successful response strategy which may involve selective scanning for known items, techniques of using partial information to guess correct answers, or efficient time use. Students possess a dichotomy with regard to their perceptions of the purposes and significance of tests (Deyhle, 1986). Acculturation has been found as an influence on both achievement and ability tests (Hoffman, et al, 1985; Guilmet, 1983). Both authors suggest that acculturation and test motivation are associated.

The second most influential factor reported in the literature is language; that is, inability of many Indian students to read the questions accurately or to be able to give appropriate verbal responses. Tests which avoid language use are not subject to underestimation as much as those that depend on verbal instructions and reading. To illustrate this the Hiskey-Nebraska test of Learning Aptitude, a non-verbal test designed for use with deaf children produced estimates of higher potential for Indian children than the Wechsler Test (Shutt, 1962). The influence of English being learned as a second language is further reinforced by the fact that many Indian first languages are unwritten.

The deficiency of the student to have the assumed experience or cognitive structure to respond to certain items is due both to the culture and to the setting in which many children are reared. The isolated, rural environment of many reservation settings, the restricted poverty of many families, and the cultural ties that promote continued identification with the tribe deny students important knowledge of the outside world. Fortune (1985) found that a majority of the Indian students in an economically deprived reservation setting lacked experience to understand the examples that teachers use in instruction and, consequently, the experience to perform on achievement tests. Studies of intelligence and aptitude test results of one tribe suggest nonstandard Indian performances (Mishra, 1981). These results are further substantiated by unique patterns of measured Indian aptitude. On the WISC-R test Indian children show a pattern of greater strength than the norm population in relational, holistic, and right hemisphere information processing (Browne, 1984). The Indian children in two reservation settings demonstrated a performance pattern on the Wechsler different from normal and learning disabled Anglo children. Spatial abilities were more well-developed than sequencing skills (McShane & Plas, 1982; Diessner & Walker, 1986).

Although Boloz and Varrati (1983) found that test scores for Indian students were higher for those who had the best attendance records and stayed in the same district, many Indian students live in discouraging situations where there is little congruence between their life experiences and the skills needed for testing. These students often do not speak English outside of school. In addition, there are few books available for them to read. Personal and community poverty, aggravated by lack of industrial development and employed role models, does little to stimulate student awareness of mathematics and its applications.

Considerations for Test Use

Applications of the principles of test theory to the sources of underestimation can lead to several useful maxims in using standardized tests with many American Indian students. Prior to the administration of

standardized tests to American Indian students or to the interpretation of American Indian test data the test user can do several things that may contribute to better test-taker performance. For ability, aptitude, and intelligence testing one should assure that the students have had exposure to the experiences assumed in the design of the test, the opportunity to develop the requisite skills, and the circumstances necessary to value a successful test performance. For achievement tests, one should make certain that the students have been instructionally exposed to the content of the test and have had opportunity to apply this content; the students have had experience in the taking of the test, are test wise and able to understand test instructions and time requirements; and the test is to be or was administered at a time similar to when it was normed and the test has Indian norms. A few of the national testing corporations such as the developers of the California Achievement Test are developing Indian norms for their tests.

Several papers offer additional reading and help in the area of testing Indian students. They include guidelines for testing bicultural children (Bernardoni, 1967) and for second language testing (Upshur and Fata, 1968), as well as annotations of tests found appropriate for use with American Indians (Educational Testing Service, 1977 and 1982).

Recommendations for future research appear fraught with problems. The natural recommendation for most tribes to become involved in the process of developing their own tests has to be considered in the light of the high costs and resource requirements needed to develop a quality test. Tribal differences and dispersion of many Indian students would suggest that tribally developed tests may lack enough general applications across tribes for merit. Further research may be better invested in documenting the similarity and differences of test reactions and in the development of intervention programs to teach test administrators to use tests in an appropriate manner with American Indian students.

REFERENCES

Aiken, Lewis R. PSYCHOLOGICAL TESTING AND ASSESSMENT. Boston: Allyn and Bacon, Inc., 1976.

Bernardoni, Louis C. "The Testing of Bicultural Children", SHARING IDEAS 4, 2 (1967): 1-5. ED 077 977.

Boloz, Sigmund A., & Varrati, Richard. "Apologize or Analyze: Measuring Academic Achievement in the Reservation School," JOURNAL OF AMERICAN INDIAN EDUCATION, 23, 1 (1983): 23-8.

Brown, Frederick G. PRINCIPLES OF EDUCATIONAL AND PSYCHOLOGICAL TESTING. New York: Holt, Rinehart and Winston, 1976.

Browne, Dauna Bell. "WISC-R Scoring Patterns among Native Americans of the Northern Plains," WHITE CLOUD JOURNAL 3, 2 (1984): 3-16.

Deyhle, Donna. "Success and Failure: A Micro-ethnographic Comparison of Navajo and Anglo Students' Perceptions of Testing," CURRICULUM INQUIRY, 16, 4 (1986): 365-89.

Dessner, Rhett, & Walker, Jacqueline L. "A Cognitive Pattern of the Yakima Indian Students," JOURNAL OF AMERICAN INDIAN EDUCATION 25, 2 (1986) 39-43.

Educational Testing Service (ETS). TESTS FOR AMERICAN INDIANS. Princeton: ETS, 1977. ED 213 546.

Educational Testing Service (ETS). TESTS FOR AMERICAN INDIANS. Princeton: ETS, 1982. ED 227 995.

Fortune, Jim C. CHOCTAW COMPREHENSIVE SCHOOL STUDY. Philadelphia, MS: Choctaw Heritage Press, 1985.

Guilmet, George M. THE INAPPROPRIATENESS OF STANDARDIZED TESTING IN A CULTURALLY HETEROGENEOUS MILIEU: A NAVAJO EXAMPLE. Los Angeles: University of California, 1983. ED 261 830.

Hoffman, Tom, et al. "Measured Acculturation and MMPI-168 Performance of Adults," JOURNAL OF CROSS-CULTURAL PSYCHOLOGY 16, 2 (1985): 243-56.

Hunter, John E. FAIRNESS OF THE GENERAL APTITUDE TEST BATTERY: ABILITY DIFFERENCES AND THEIR IMPACT ON MINORITY HIRING RATES. Uses Test Research Report No. 46. Sacramento: California State Department of Employment Development, 1983. ED 237 534.

McDiarmid, G. L. THE HAZARDS OF TESTING INDIAN CHILDREN. Las Cruces: ERIC/CRESS, 1972. ED 055 692.

McShane, Damian Anthony, & Plas, Jeanne M. "Wechsler Scale Performance Patterns of American Indian Children." PSYCHOLOGY IN THE SCHOOLS 19, 1 (1982): 8-17.

Mishra, Shitala P. "Relationship of WISC-R Factor Scores to Academic Achievement and Classroom Behaviors of Native American Navajos." MEASUREMENT AND EVALUATION IN GUIDANCE 14, 1 (1981): 26-30.

Neely, Renee, & Shaughnessy, Michael F. ASSESSMENTS AND THE NATIVE AMERICAN. Portales: Eastern New Mexico University Psychology Department, 1984. ED 273 889.

Shutt, Darold L. FAMILY PARTICIPATION IN THE PSYCHOLOGICAL EVALUATION OF MINORITY CHILDREN. Paper presented to Southwestern Orthopsychological Association Meeting, Galveston, TX, November, 1972. ED 071 830.

Upshur, John A., & Fata, Julia, (Eds). PROBLEMS IN FOREIGN LANGUAGE TESTING: PROCEEDINGS OF A CONFERENCE HELD AT THE UNIVERSITY OF MICHIGAN, SEPTEMBER, 1967. Ann Arbor, MI: University of Michigan Research Club in Language Learning, 1968. ED 022 162.

Prepared by:

William Brescia
Mississippi Band of Choctaw Indians
Ph.D. Candidate
University of Indiana
Bloomington, IN

Jim C. Fortune
Virginia Polytechnic Institute
Blacksburg, VA

March 1988

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI) U.S. Department of Education, under contract number 400-86-0024. Its contents do not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. government.

For further information contact: ERIC/CRESS, Dept. 3AP, Box 30001, New Mexico State University, Las Cruces, N.M. 88003-0001 (505) 646-2623